

Tendance de la visualization en France

2019

GdR IG-RV Visualization 2019.

Ce document a été partiellement rédigé à partir de la synthèse des réflexions produites lors d'un panel qui a eu lieu à la journée visu 2019. Ces réflexions ont été transcrites puis librement éditées afin de rentrer dans un format court.

Les panelistes présents été : Thomas Baudel, [IBM](#); Caroline Goulard, [Dataveyes](#); Petra Isenberg, [Inria](#); Julien Tierny, [CNRS](#); Romain Vuillemot, [Ecole Centrale de Lyon](#).

Editeurs: Samuel Huron [Telecom Paris](#), Nadia Boukhelifa [INRA](#).

La communauté de recherche dans le domaine de la visualisation est vivace et en pleine extension. On a assisté à la multiplication de conférences internationales durant ces 5 dernières années. Par exemple [OpenVis](#) dont l'édition 2018 a eu lieu à Paris ainsi que l'exposition [123 Data](#), ou encore [Information Plus](#) dont la 2nde édition 2018 a eu lieu à Berlin, et aussi des prix comme [information is beautiful award](#) qui a eu lieu à Londres. Ces événements, qui viennent s'ajouter aux rendez-vous historiques comme [IEEE VIS](#), [EUROVIS](#), [Pacific VIS](#) et [Journée Visu](#), montrent la vivacité et la pluridisciplinarité du domaine. Ci-dessous nous relatons les avancées, questions, enjeux récents, ainsi que la place de la France dans le domaine.

1 # Les avancées récentes (mais aussi les domaines en perte de vitesse, voire épuisés) ?

Alors que pendant les 20 dernières années, la communauté de visualisation a produit beaucoup de contributions dans la création de nouvelles représentations de données et techniques d'interaction pour des experts sur des ordinateurs personnels, on peut voir que ces derniers temps la communauté se diversifie à plusieurs niveaux : au niveau des sources de données, supports et dispositifs d'affichage, méthodes d'analyse et types d'audience.

Une communauté réduite pour SciVis, mais convergence vers des disciplines connexes (JT,PI): la communauté de la visualisation scientifique traditionnelle a connu un déclin ces dernières années. Après avoir connu un essor rapide et important au début des années 90, où de nombreux problèmes clés ont été étudiés, la communauté semble avoir atteint une certaine maturité sur les problèmes accessibles, tandis que les problèmes ouverts et non résolus s'avèrent de plus en plus difficiles. En termes de nouvelles tendances pour SciViz, il existe un intérêt croissant pour les données d'ensemble, multimodales et abstraites (ex: nuages de points en grande dimension); mais aussi une évolution vers d'autres domaines connexes tels que la visualisation d'informations et l'apprentissage automatique.

Un impact externe en recul où l'analyse automatique occupe le devant de la scène (TB): on considère que l'impact réel de la communauté de recherche peut être mesuré par son impact sur le monde extérieur. L'âge d'or du domaine pour l'analyse visuelle des données, a commencé avec une société comme Spotfire (1995), s'est prolongé dans les années 2000 avec Tableau, Qlik et BusinessObjects. Puis, il y a eu un déclin à partir de 2010, lorsqu'avec des outils d'analyse tels que R, scikitlearn et la nouvelle vague d'IA (TensorFlow), l'analyse automatique a pris la place centrale.

Cependant les nouveaux enjeux propre aux interactions humain-IA et société-IA tel que le pilotage d'IA, l'explicabilité des résultats produisent aussi de nouvelles opportunités de recherche et des besoins pour les techniques de visualisation. Les approches visuelles et interactives en IA sont un champs de recherche plus en plus à la mode. L'enjeu étant de trouver comment IA et humains se complètent dans la prise de décision utilisant des modèles d'apprentissage automatique.

Le panel a identifié deux thèmes clés comme sujets d'actualité, à savoir:

Apprentissage automatique avec l'humain dans la boucle (CG,TB): de nombreux travaux publiés se situent maintenant à l'intersection de l'IA et de la visualisation. Comme les modèles d'apprentissage automatique servent en outils d'aide à la décision, la communauté de la visualisation s'intéresse maintenant à la façon de faire participer les humains à la construction d'algorithmes plus puissants; l'audit des algorithmes, et en ajustant ces algorithmes pour fournir des résultats pertinents pour les humains, en particulier lorsqu'il s'agit de jeux de données bruités.

Les framework pour décrire le design de visualisation sont de plus en plus mature (RV): les cadres conceptuels et langages formels spécifiant des visualisations (par exemple, Vega, Vega Lite, d3.js) sont de plus en plus populaires, utilisés et adoptés. En effet, l'idée de disposer d'un langage commun pour décrire la visualisation est attrayant pour plusieurs raisons. De nombreux outils reposent désormais sur des frameworks tels que Vega, ce qui facilite l'interaction entre ces outils. La combinaison de ces cadres avec des outils de "sketching", par exemple, peut ouvrir le champ à d'autres personnes intéressées par la conception visuelle mais pas nécessairement par les détails de mise en œuvre ou de programmation.

2 # Les questions actuelles et les principaux verrous identifiés?

Données multi-échantillons et calcul haute performance (JT): le premier défi est lié à plusieurs ensembles de données, tels que ceux issus de la simulation multi-physique. Il est difficile de trouver des tendances et des schémas globaux décrivant avec précision le processus physique sous-jacent lors de l'utilisation de sources de données disparates. Certaines des questions de recherche liées à ce défi chevauchent des problèmes étudiés par la communauté ML (machine learning). Le deuxième grand défi concerne le calcul haute performance, la mise en parallèle des algorithmes existants et le calcul in-situ et progressif.

Alphabétisation en visualisation (PI): alors que de nombreux outils et techniques de visualisation sont maintenant disponibles pour visualiser des données, on ne sait pas encore comment le grand public comprend et reçoit ces visualisations. Nous devons réfléchir aux conséquences des types de données et des types de visualisation que nous exposons en tant que praticiens de la visualisation ou chercheurs.

Qualité des données (RV): les données sont générées par l'homme et sont imparfaites (erreurs, valeurs manquantes, valeurs aberrantes, etc.). Nous avons besoin d'outils pour aider à nettoyer ces ensembles de données avant de les introduire dans nos modèles d'apprentissage. De plus, nous devons aider les humains dans ce processus de nettoyage, en particulier pour les jeux de données temporelles.

Intégrer l'édition des données et la visualisation des données (TB): il s'agirait d'un objectif à long terme. La plupart des travaux d'analyse de données ont lieu avant la modélisation, l'exécution d'algorithmes et la communication des résultats. De nombreux efforts humains sont consacrés au nettoyage et à la restructuration des données, ainsi qu'à une analyse exploratoire des données. Nous devons soutenir l'ensemble du processus d'analyse des données en prenant en compte tous les aspects cette analyse.

3 # Les enjeux de la discipline et quelle est sa place au sein de la société?

Dans cette section on présente deux enjeux majeurs pour la société, des enjeux pour la discipline et un enjeu commun.

Pour la société : Digital Literacy, explanation and democratization

La visualisation d'information peut être un outil majeur pour aider les citoyens à mieux comprendre le monde numérique, à expliquer et à démocratiser des domaines très techniques, scientifiques ou légaux. Parmi ces domaines on peut citer par exemple l'alphabétisation aux données, "l'intelligence artificielle", la gestion de la vie privée dans le numérique (CG), mais aussi une meilleure compréhension des traitements algorithmiques (CG, TB). Être capable d'expliquer simplement aux citoyens des processus complexes et un des défis scientifique majeurs pour l'humanité dans laquelle le domaine de la visualisation peut contribuer.

Pour la société : Visualization Science for the masses (JT)

Un problème rencontré par des scientifiques dans de multiples domaines, et le manque de confiance du grand public dans les données et résultats des travaux scientifiques. La visualisation des données peut être un outils très puissant pour résoudre ce problème. Comment améliorer la visualisation des données scientifiques pour les citoyens?

Pour la société : Visualization as an exploration tool

Trop souvent la visualisation est seulement perçue comme un outil de communication. Un des défis majeurs dans la société est d'utiliser la visualisation non pas uniquement comme un outil de communication, mais comme un outil d'exploration. Nous devons faire un meilleur travail pour enseigner cela. (RV)

Pour la discipline : Recrutement, décentralisation, organisation et financement

Si on compare le domaine de la visualisation en France et en Allemagne qui sont des pays avec des populations équivalentes, l'Allemagne a beaucoup plus de professeurs dans le domaine ce qui amène à plus de publications internationales et une plus forte présence (PI). Par ailleurs, on remarque qu'en France, la communauté semble plus concentrée sur quelques institutions géographiquement proche, là où en Allemagne, les professeurs sont plus distribués dans différents lieux et semble mieux organisés. Cela pose des problèmes à la discipline, par exemple pour attirer des financements (avoir des relecteurs compétents), la visibilité du domaine, le recrutement d'étudiants compétents. Donc les défis sont comment mieux s'organiser (TB), recruter plus (PI), décentraliser (PI) et mieux financer (TB, PI).

Pour la discipline et la société : Enseignement de la visualisation plus tôt (CG, RV)

A l'école, on enseigne à lire des cartes géographiques mais trop rarement des diagrammes mathématiques. Enseigner la visualisation plus tôt dans les cursus scolaire est un enjeu important afin que tout citoyen puisse mieux lire et comprendre des représentations graphiques de données. Ce type d'enseignement pourrait aussi être un outil efficace pour l'alphabétisation numérique et l'alphabétisation des données.

4 # La place de la recherche française au niveau international?

D'après une étude menée par Petra Isenberg et al., de 1990 à 2018, la France est le 8ème pays publiant le plus de papiers dans la revue IEEE TVCG derrière les US, l'Allemagne, le Canada, l'Autriche, la Chine et l'Angleterre. À l'échelle de l'Europe, elle est 4ème avec un peu plus d'une centaine de publications sur cette période de 28 ans contre 400 pour l'Allemagne (1er) et moins d'une dizaine pour le Portugal (dernier). L'activité de recherche française dans cette conférence est dominée par l'INRIA, le CNRS, et l'université Paris Sud.

La présence dans les comités de programme l'année dernière pour IEEE VIS 2018, est de 4 chercheurs Français dans le comité de programme *Infovis*, et 4 autres travaillant à l'étranger mais provenant de laboratoires français (4/63), dans le comité de programme de visualisation scientifique, 1 français sur 52. Les français sont aussi présents en demi teinte dans les conférences émergentes par exemple *Open Viz* a été organisé par l'EM Lyon, alors que *Information plus* organisé cette année à Berlin et *information is beautiful award* ne compte aucun chercheur français dans l'organisation. A Paris une exposition artistique (1,2,3 Data) a été organisée à la fondation Electra par David Bihanic (Maître de conférence à la Sorbonne).

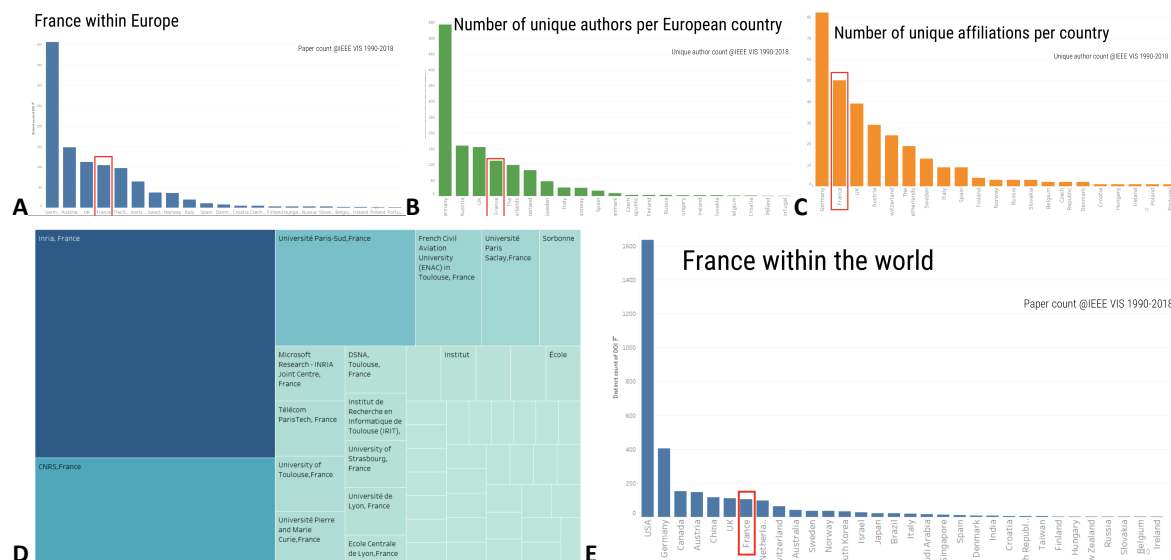


Figure 1 . Visualization extraite des slides de Petra Isenberg sur le nombre de publications par pays dans la revue IEEE Transactions on Visualization and Computer Graphics. A) Nombre de publication par pays européen, en rouge la france, B) Nombre d'auteurs uniques par pays européens, C) Nombre d'affiliation unique par pays. D) Répartition des publication par affiliation en France, E) Nombre de publication par pays.